Handwritten Character Recognition Using Soft Computing Technique

Prashanth G K[1], M A Jayaram[2], Navya M C[3], Nanda V H[4]
[1, 2, 3, 4] Dept of Master of Computer Applications
[1, 2, 3, 4] Siddaganga Institute of Technology, Tumkur, Karnataka, India

Abstract- *Handwritten digit recognition has recently been of very interest among the researchers because of the evolution of various Machine Learning techniques. In this article, we first collect the handwritten characters and build a database of characters, the features dataset, the collected dataset is used for KNN to training, verify and validate the dataset with the known and unknown by the KNN network, finally, the results of both dataset resulted in an accuracy of 99.66 and 99.33 respectively*

Keywords- kannada and english dataset, KNN

I. INTRODUCTION

In recent years, automatic processing in native languages is becoming more popular. Online Handwriting Recognition (OHR) is one such area which provides the natural way of communicating with machines using in the local language with the digital pen or hand written. Online or Handwriting Recognition gaining more attention due to the growing attractiveness of handheld computers, PDA and smart phones. The challenge with the pen enabled PC is to recognize user handwritings. There are basically two categories, namely offline and Online Handwriting Recognition (OHR). In offline systems, written data are scanned and converted into image and recognition is carried out. In online systems, handwriting data is acquired while writing, which can be regarded as a dynamic representation of handwriting [1].

A survey on OHR [2] reports, work carried out in English ([3], [4]), Chinese [5] and Japanese [6] using time zone direction, chain codes, etc. For the Indian languages, the combination of Hidden Markov Model and Nearest Neighbor classifiers are used in [7] for Devanagari script is reported in 2000. Principal Component Analysis on Tamil characters is described in [8]. The elastic matching method was implemented by Niranjan et al. in 2004 [9]. In Tamil [10], pen strokes are denoted as strings of shape features and string search method is used to test stroke. Time and frequency domain features were extracted and testes with HMM [11] . Kunte [12] reported first research in Kannada OHR using a neural network as the classifier. Prasad et.al in 2009 [13] applied divide and conquer technique using 295 classes. Online Kannada Akshara recognition [14] is carried out using Statistical Dynamic Space Warping (SDSW). Ramapalli [15] implemented fusion of offline and online character recognition for better recognition efficiency of 89.7%. Venkatesh and Ramakrishnan in 2011 [16] proposed a dexterous technique for Kannada Akshara in which two set of classifier namely the primary and secondary classifier are implemented chosen and accuracy is improved from 77% to 92 %.

Many different types of classifier algorithms were implemented and tested for OHR and promising results were reported, however, meticulous experimental assessment of different kinds of classifier algorithms on a given dataset is not reported.

This paper is organized as follows. Section II briefly introduces Kannada Script and data collection methods.

The proposed methodology is presented in Section III. Section IV discusses the various results. Finally, conclusions are drawn in Section V.

II. KANNADA AND ENGLISH SCRIPT AND DATA COLLECTION

i. Kannada Script and Data Collection

Kannada is derived from the Dravidian language and is one among the twenty-two official languages recognized by the Indian Constitution. Kannada is the state language of Karnataka and spoken by around fifty million people in and out of Karnataka state. The script of Kannada language is syllabic and is called “Kannada script” which is evolved from “Kadamba script” [12] and written from left to right. It has sixteen vowels and thirty-four consonants. The character set of Kannada script is given in Figure 1.

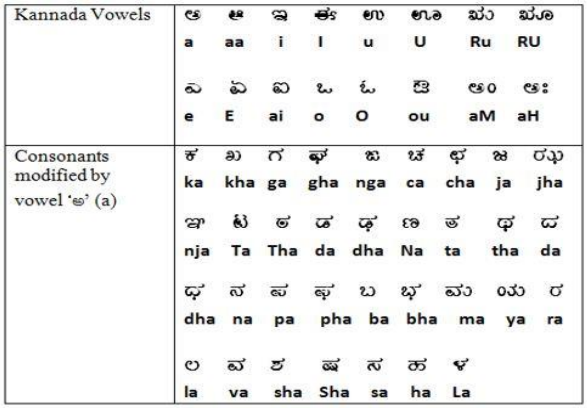


Figure-1 Kannada script

ii. English Script and Data Collection

The modern English alphabet is a Latin alphabet consisting of 26 letters, each having an upper- and lower-case form. It originated around the 7th century from the Latin script. Since then, letters have been added or removed to give the current Modern English alphabet of 26 letters (the same as in the ISO basic Latin alphabet). Shown in figure 2

Figure-2 English Script

iii. Processing of Handwritten Data

The English and kannada letter, numbers were used in the work, first the letters were hand written with 10 rows and 10 column and the written character were scanned for processing, the single characters were segmented and stored separately for the next stage for feature extraction the sample handwritten and individual character is shown in below Table 1 and Table 2.

Table -1 Hand written and scanned image

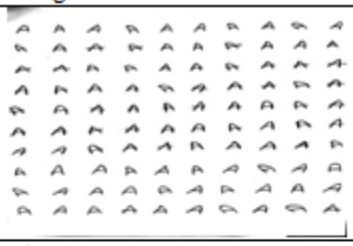
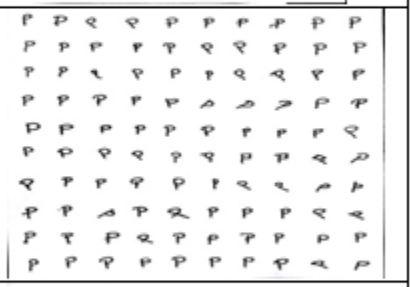
Sl.no	letter	Images hand written
1.	A	
2.	p	
3.	Aha	

Table -2 Separated Handwritten Character

Sl.no	letter	Separated Characters
1.	A	
2.	p	
3.	Aha	

III. METHODOLOGY

The 9800 sample of Kannada and English characters were used in the work, the 1200 samples were new to the dataset, and the total samples of hand written were 11,000

were used. The figure 3 depicts the dataset used in the work include the Kannada & English letters. The Kannada & English handwritten character from the database is processed, dominant features are extracted and stored in database, the steps involved is represented in the figure 4.

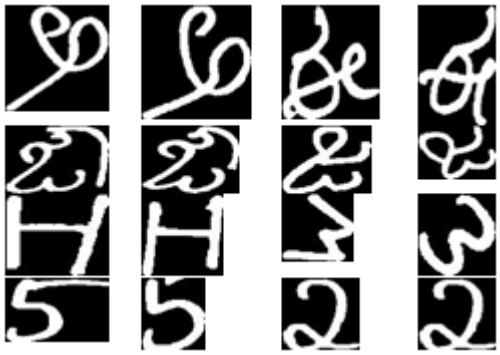


Figure-3 Database Kannada and English set

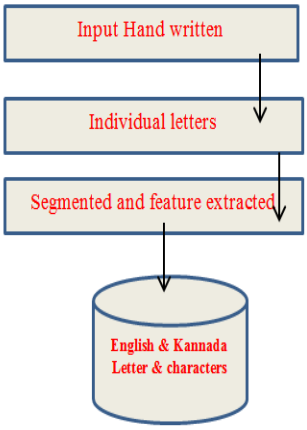


Figure-4 Flow diagram for feature extraction and storage

In the next stage the dataset is arranged/ordered using supervised learning methodology for grouping, the letter of both used, later there were used for training using feed forward neural network the network with the accuracy of 99.6% were found, the entire stage is shown in figure 5. The database is now with the grouped one.

Figure-5 Dataset trained using FFNN

Once the network is ready, the next stage is the testing network, the entire processes involved in testing with the old dataset and the new dataset with query handwritten images neither used in the processes. The system performed well with both the set identifying the Kannada or English letters separately, the entire stage is shown in the below figure 6.

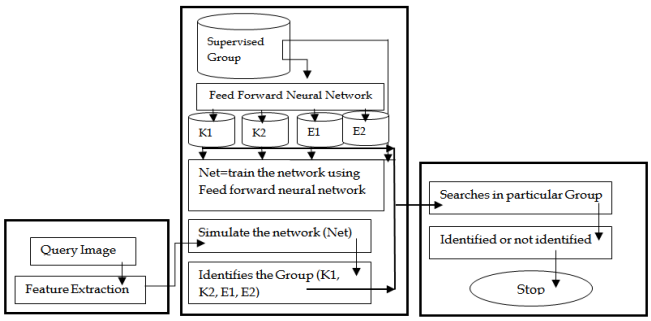


Figure-6 Flow diagram involved for network validation

The system is tested with other images that are not handwritten, the system at the beginning will give the message or result stating that the image is not a hand written images and the system itself won't enable the other related process.

IV. RESULTS AND DISCUSSION

To begin with the user interface the figure 7, the main form is loaded and the next stage is to click start, it allows to load the image, the feature are extracted from the menu driven in the processes the system allows to identify either the loaded image or character is digit or letter Kannada or English. In the next stage moves in the direction and identifies the letter of digit and gives the result.

Figure 8 through figure 10 all the loaded character are identified correctly, from the figure 11 through figure 13 there is mismatch in recognition of the letter of digit, this is the wrongly classification here the system fails to identify the character, it might be the hand written letter may be similar in the writing.

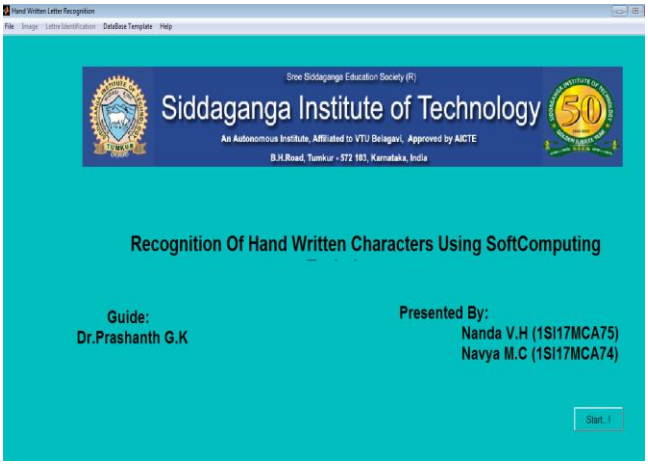


Figure-7 Startup Screen

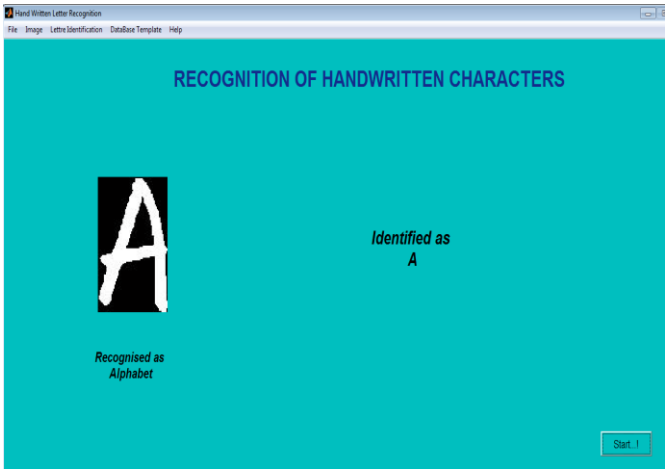


Figure-8 Correctly Identified English Letter

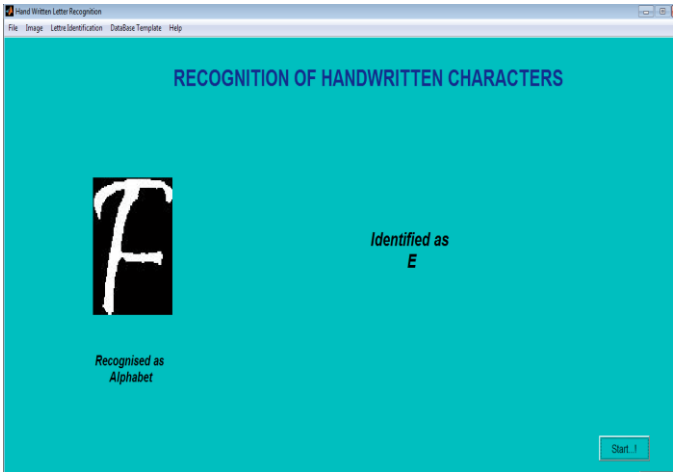


Figure-11 Mismatch Identified Letter-1

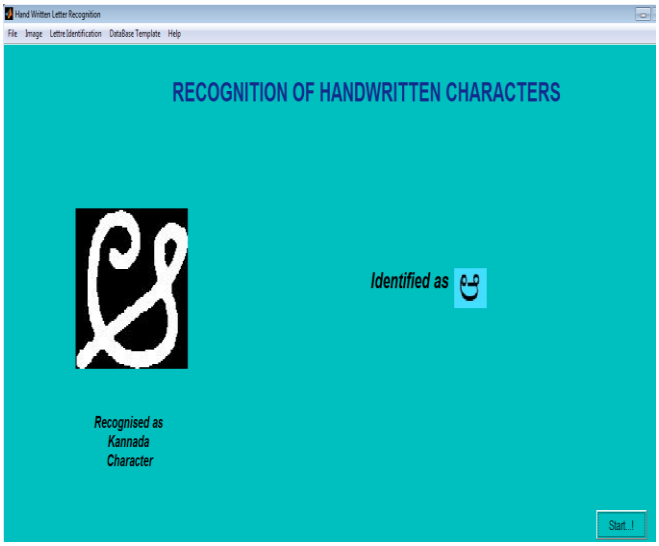


Figure-9 Correctly Identified Kannada Letter

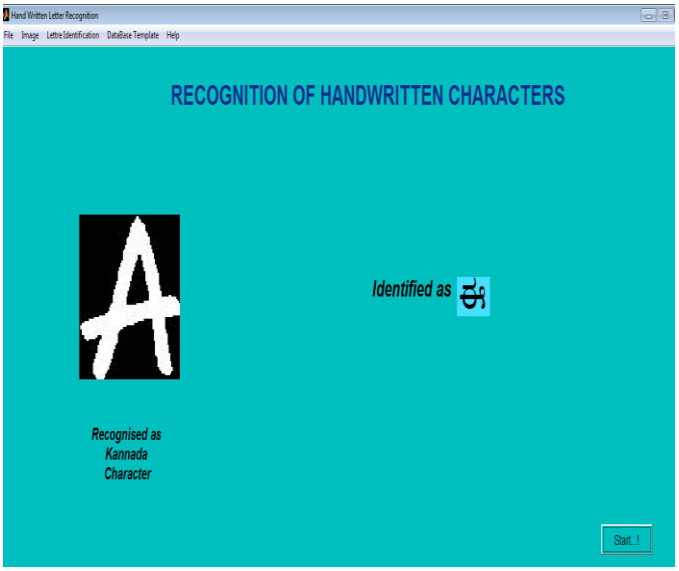


Figure-12 Mismatch Identified Letter-2

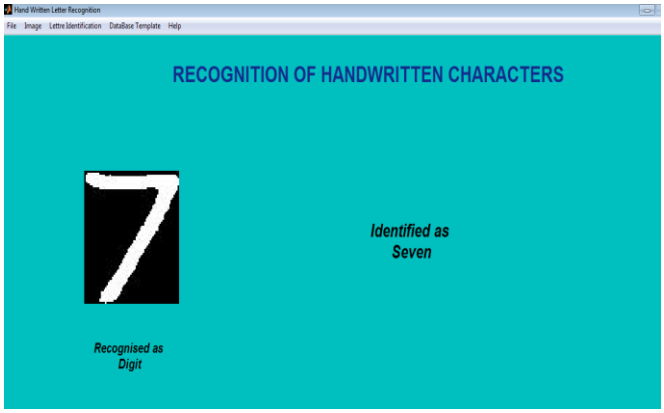


Figure-10 Correctly Identified Number

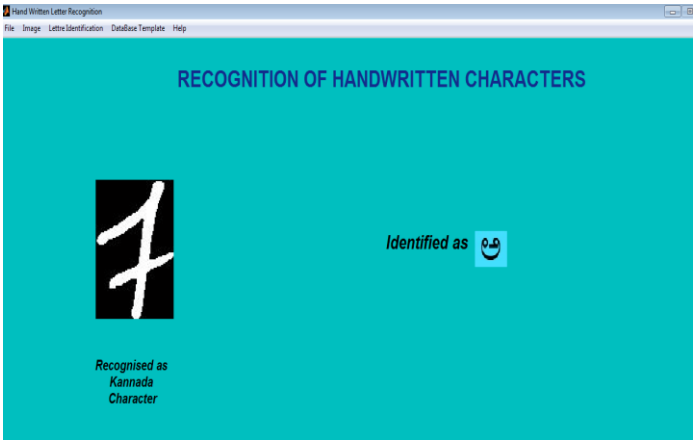


Figure-13 Mismatch Identified Letter-3

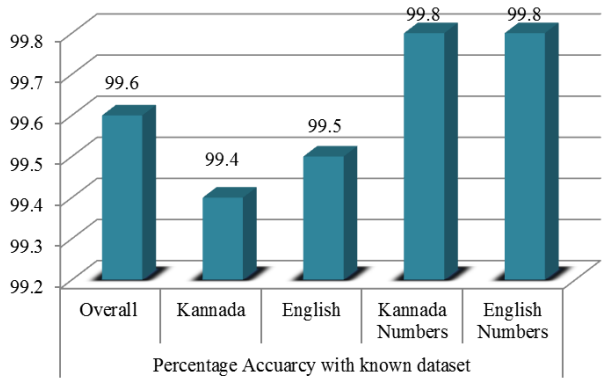


Figure-14 Overall percentage-1

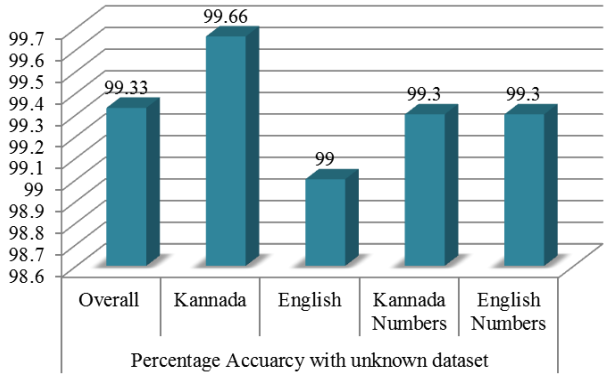


Figure-15 Overall percentage21

The figure 14 shows the overall percentage of identification with the known dataset with in the given group, but the system gave good results with the unknown dataset that is not present, because the handwritten characters were not used in training, these used only for validation the system is signified in figure 15.

V. CONCLUSION AND FUTURE SCOPE

In this paper, we have used the concept of KNN to build and identify the letter or characters, a very high amount of accuracy can be achieved, we are able to get an accuracy of 99.66% of overall dataset with the known dataset. The same network was used for identifying the Kannada and English characters separately, with the accuracy of 99.4%, 99.5% with known grouping respectively, the other part with the accuracy of 99.66%and 99% achieved for Kannada and English letter respectively.

VI. ACKNOWLEDGMENT

The authors would like to acknowledge the research laboratory provision rendered by Department of Master of Computer Applications, Siddaganga Institute of Technology, Tumakuru, and Karnataka India in this research work.

REFERENCES

[1] S. Ramya, Kumara Shama, Unrestricted Online Kannada Handwriting Recognition Using Data Mining Techniques, Jour of Adv Research in Dynamical & Control Systems, 13-Special Issue, 2017.

[2] Plamondon, R. and Srihari, S.N. On-line and offline handwriting recognition: a comprehensive survey. IEEE Trans. Pattern Anal. Mach. Intel. 22 (2000) 63–84.

[3] Jaeger, S., Manke, S., Reichert, J. and Waibel, A. Online handwriting Recognition : the NPen ++ recognizer. International Journal on Document Analysis and Recognition 3 (2001) 169–180.

[4] Bahlmann, C., Haasdonk, B. and Burkhardt, H. Online handwriting Recognition with support vector machines-a kernel approach. Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition, 2002, 49-54.

[5] Jaeger S., Liu, C.L. and Nakagawa, M. The state of the art in Japanese on-line handwriting recognition compared to techniques in western handwriting recognition. IJDAR 6 (2003) 75-88.

[6] Liu, C.L., Jaeger, S. and Nakagawa, M. Online recognition of Chinese characters: the state-of-the-art. IEEE Trans. on Pattern Analysis and Machine Intelligence (2004) 198-213.

[7] Connell, S.D., Sinha, R.M.K. and Jain, A.K. Recognition of Unconstrained On-Line Devanagari Characters. 15th IEEE International Conference on Pattern Recognition, 2000, 368–371.

[8] Deepu, V., Madhvanath, S. and Ramakrishnan, A.G., 2004, August. Principal component analysis for online handwritten character recognition. In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on(Vol. 2, pp. 327-330). IEEE.

[9] Joshi, N., Sita, G., Ramakrishnan, A.G. and Madhvanath, S. Comparison of elastic matching algorithms for online Tamil handwritten character recognition. Ninth International Workshop on Frontiers in Handwriting Recognition, 2004, 444-449.

[10] Aparna, K.H., Subramanian, V., Kasirajan, M., Prakash, G.V., Chakravarthy, V.S. and Madhvanath, S. Online Handwriting Recognition for Tamil. 9th International Workshop on Frontiers in Handwriting Recognition, 2004.

[11] Babu, V.J., Prasanth, L., Sharma, R.R., Rao, G.V.P. and Bharath, A. HMM-Based Online Handwriting Recognition System for Telugu Symbols. 9th International Conference on Document Analysis and Recognition, 2007.

[12] Kunte, S.R. and Samuel, S. Wavelet features based online recognition of handwritten Kannada characters. Jl Visualization Society of Japan (2000).
[13] Mahadeva Prasad, M., Sukumar, M. and Ramakrishnan, A.G. Divide and Conquer Technique in Online Handwritten Kannada Character Recognition. Proc. Intl. Workshop on Multilingual OCR, 2009.
[14] Rituraj, K., Mohan, P., Shashikiran, K. and Ramakrishnan, A.G. Unrestricted Kannada Online Handwritten Akshara Recognition using SDTW. Journal of Universal Computer Science 17 (1) (2011) 94-106.
[15] Rampalli, R. and Ramakrishnan, A.G. Fusion of Complementary Online and O ine Strategies for Recognition of Handwritten Kannada Characters. Journal of Universal Computer Science 17 (1) (2011) 81- 93.
[16] Murthy, V.N. and Ramakrishnan, A.G. Choice of Classifiers in Hierarchical Recognition of Online Handwritten Kannada and Tamil Aksharas. J. UCS 17 (1) (2011) 94-106.